공간 데이타베이스 시스템에서 순환 속성을 지원하는 공간색인구조의 성능평가

김홍기[†] · 선휘준[††]

요 약

공간 데이타베이스 시스템의 성능을 향상시키기 위해서는 동적 및 정적 환경에서 발생하는 공간객체를 효율적으로 관리할 수 있는 공간색인방법이 필요하다. 그리고 검색의 성능을 높이기 위해서는 공간국부성을 고려한 공간색인방법이 요구되며, 공간국부성은 객체들의 위치 속성과 관계가 있다. 기존의 공간색인방법들에서는 객체가 가질 수 있는 순환적인 위치 속성이 고려되지 않았다.

본 논문에서는 순환 및 선형 도메인들로 구성된 검색공간에서 공간적으로 이웃하는 객체들을 결집시키기 위한 공간색인구조인 CR-트리를 소개한다. 그리고 객체의 순환적인 위치 속성을 고려한 공간색인구조를 사용함으로써 적중률 및 버켓이용률을 향상시킬 수 있음을 실험을 통하여 보인다.

Performance Evaluation of a Spatial Index Structure Supporting the Circular Property in Spatial Database Systems

Hong-Ki Kim[†] and Hwi-Joon Seon[††]

ABSTRACT

In order to increase the performance of spatial database systems, a spatial indexing method is necessary to manage spatial objects efficiently in both dynamic and static environments. A spatial indexing method considering a spatial locality is required to increase the retrieval performance. And the spatial locality is related to the location property of objects. The previous spatial indexing methods did not consider the circular location property of objects.

In this paper, we introduce the CR-Tree that is a spatial index structure for clustering spatially adjacent objects in which a search space is constructed with the circular and linear domains. Using a spatial index structure considered a circular location property of objects, we show that high hit ratio and bucket utilization are increased through the simulation.

1. 서 론

멀티미디어 시스템과 같은 최근의 응용분야에서는 더욱 복잡하고 많은 양의 공간적 성질을 갖는 데이타를 취급하고 있음으로, 이러한 데이타를 효율적으로 관리하기 위한 공간 데이타베이스 시스템이 필요하다[3,4,7,8].

공간 데이타베이스 시스템에서 공간객체(이하 객체라 함)를 관리하기 위해서는 많은 저장공간과 처리시간이 요구되며, 적절한 색인구조와 질의처리 기법을 적용함으로써 공간 질의를 효율적으로 처리할 수 있다[1,2,5,9].

기존의 공간색인방법들이 가정하는 검색공간에서 도메인은 객체의 속성이 가질 수 있는 최소와 최대값 사이의 선형적 순서 범위로 구성된다. 그러나 멀티미디어 시스템에서 취급되는 객체 또는 시간의 개념이 포함되어 있는 객체 등은 순환적인 속성을 가질 수 있으며, 이러한 객체의 속성을 효율적으로

† 동신대학교 컴퓨터학과 부교수
†† 서남대학교 컴퓨터영상 · 정보통신학과 조교수

반영하기 위해서는 순환적인 성질을 갖는 도메인으로 구성된 검색공간이 요구된다.

다음은 도메인이 각각 0°에서 360°, 0시에서 24시인 순환적 환경에서 발생하는 영역질의와 최대근접질의의 예이다.

i) "X가 350°부터 20°사이이고 Y가 15°부터 30° 사이에 있는 객체들을 검색하라"

ii) "23시에 가장 근접한 시간에 발생된 객체를 검색하라"

i)의 영역질의는 순환적인 성질이 포함된 질의범위와 순환적이지 않는 질의범위로 이루어진 예이며, ii)는 근접을 반드시 순환적인 환경에서 고려해야 하는 경우이다. 순환적인 환경이 주어졌을 때 기존의 공간색인방법들에서는 인접한 객체들이 서로 다른 버켓에 분산되어 저장됨으로 예에서와 같은 질의 형태에 대하여 효율적인 검색이 이루어지지 못한다.

[11]에서는 객체의 순환적인 위치 속성을 고려하기 위해 순환 도메인(circular domain)과 순환 영역 질의(circular region query)를 정의하였으며, 새로운 동적색인구조인 PR-화일(Pointed Region-File)을 제안하고 그 특성을 실험을 통하여 보였다. 그러나 실험에서는 균일 및 군집분포를 갖는 점 객체만을 데이타로 취급하였다.

본 논문은 크게 두 가지 영역으로 구분할 수 있다. 첫 번째는 순환적인 성질을 갖는 도메인이 포함된 검색공간에서 공간국부성(spatial locality)을 고려하여 객체를 결집시키는 공간색인구조인 CR-트리를 소개하는 것이다. CR-트리는 PR-화일을 확장한 구조로 점 및 영역객체를 모두 취급할 수 있다.

두 번째 연구는 순환적인 위치 속성을 갖는 객체의 저장 및 검색에 따른 성능을 평가하는 것이다. 이를 위해 색인 처리기로는 CR-트리를 이용하며, 성능은 적중률, 디스크 접근 횟수, 버켓이용률 그리고 색인 구축 및 질의 처리 시간 등의 성능 평가 지수에 따라 비교 분석된다.

2. CR-트리

공간국부성을 고려하여 객체를 결집시키기 위해서는 전체 색인의 재구축 없이 부검색공간의 영역 경계를 변경할 수 있는 방법이 필요하다. 또한 질의의 성능을 높이기 위해서는 빈 공간이 최소가 되는 부검색공간을 형성하여야 한다. 이 장에서는 순환 및 선형 도메인들로 구성된 검색공간에서 공간적으로 이웃하는 객체들을 결집시키기 위한 공간색인구조인 CR-트리의 구조 및 알고리즘을 기술한다.

2.1 CR-트리의 구조

CR-트리(Circular R-Tree)는 R-트리와 유사한 색인구조를 가지며, 객체들이 저장되는 버켓과 버켓을 접근하기 위해 경로 정보를 제공하는 노드로 구성된 높이 균형 트리이다[10].

CR-트리에서 N차원 검색공간은 S개의 순환 도메인과 $N-S$개의 선형 도메인(linear domain)으로 구성된다(단, $0 \le S \le N$). 만약 S가 0이면, 기존의 공간색인방법들이 전제로한 선형 도메인으로만 구성된 검색공간이다.

CR-트리의 버켓에는 헤더와 하나 이상의 객체가 저장되며, 검색을 위해 각 객체에는 영역값 R이 할당된다. 하나의 객체는 다음과 같이 표현된다.

$$(R,\ \text{객체})$$

R은 해당 객체가 검색공간에 차지하는 영역을 N차원 최소 경계 사각형으로 근사시킴으로써 구해진다.

$$R = (R_0, R_1, \cdots, R_{N-1})$$

여기에서 R_i ($i=0,1,\cdots,N-1$)는 객체가 차원 i에서 차지하고 있는 범위를 나타내는 폐쇄 경계 구간(closed bounded interval)이다. 즉, RI_S를 폐쇄 경계 구간의 시작값, RI_E를 구간의 끝값이라고 했을 때 하나의 객체가 차지하는 부검색공간의 영역은 다음과 같이 카테시언 곱으로 나타낼 수 있다.

$$[RI_S, RI_E]_0 \times [RI_S, RI_E]_1 \times \cdots \times [RI_S, RI_E]_{N-1}$$

CR-트리의 노드는 헤더와 다음과 같은 엔트리들로 구성된 일차원 배열 구조를 갖는다.

$$(R,\ P_{child})$$

위에서 기술된 바와 같이 R은 각 차원의 폐쇄 경계 구간들로 이루어진 N차원 부검색공간이다. 그리고 P_{child}는 하위 레벨의 노드 또는 버켓의 주소지시자이다.

엔트리가 트리의 리프 노드에 존재한다면, R은 하나의 버켓이 차지하는 부검색공간을 나타내며 P_{child}는 해당 버켓의 주소지시자이다. 엔트리가 리프 노드에 존재하지 않는다면, R은 자식 노드에 속한 모든 객체를 포함하는 부검색공간이며 P_{child}는 자식 노드의 주소지시자이다.

그림 1과 그림 2는 두 개의 순환도메인으로 구성된 검색공간에서 노드의 최대 엔트리 수가 4일 때, CR-트리에서 버켓 및 부검색공간들의 포함 관계와 그에 따른 색인구조를 나타낸 것이다. 그림 1에서 R1과 R2는 순환적 성질을 갖는 부검색공간이다.

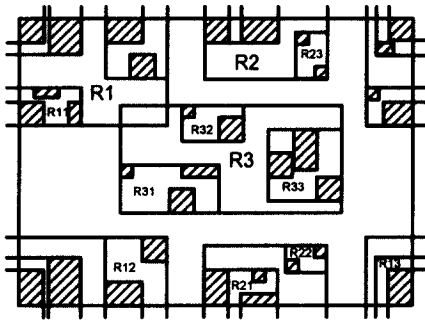



그림 1. CR-트리의 검색공간

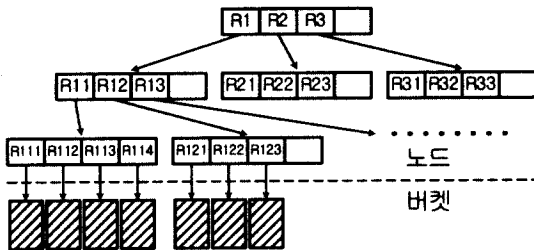



그림 2. CR-트리의 색인구조

하나의 노드에 저장할 수 있는 최대 엔트리 수를 M이라 하고, $m(\geq \lceil \frac{M}{2} \rceil)$을 최소 엔트리 수라 하자. 그리고 하나의 버켓에 저장할 수 있는 최대 객체 수를 L이라 하고, $l(\geq \lceil \frac{L}{2} \rceil)$을 최소 객체 수라고 했을 때, CR-트리는 다음과 같은 성질을 만족한다.

- 모든 리프 노드는 해당 노드가 루트가 아닐 때 최소 m, 최대 M개의 엔트리를 갖는다. 여기에서 엔트리는 버켓을 가리키는 주소지시자가 포함된 엔트리를 의미한다.
- 리프 노드의 엔트리 (R, P_{child})에서, P_{child}가 가리키는 버켓 내의 객체들이 순환적으로 결집되었다면 R은 $RI_S > RI_E$인 관계를 가지는 폐쇄 경계 구간을 포함한다.
- 루트와 리프가 아닌 모든 중간 노드는 최소 m, 최대 M개의 엔트리를 갖는다. 여기에서 엔트리는 자식 노드를 가리키는 주소지시자가 포함된 엔트리를 의미한다.
- 리프가 아닌 노드의 엔트리 (R, P_{child})에서, P_{child}가 가리키는 자식 노드의 엔트리들이 순환적으로 결집되었다면 R은 $RI_S > RI_E$인 관계를 가지는 폐쇄 경계 구간을 포함한다.
- 리프가 아닌 루트 노드는 적어도 두 개의 자식 노드를 갖는다.
- 모든 리프 노드는 동일한 레벨에 존재한다.
- 버켓이 하나 이상이면 각 버켓은 최소 l, 최대 L개의 객체를 갖는다.

[10]에서는 CR-트리에서 n개의 객체를 저장하기 위한 트리의 최대 높이 h는 $\lceil \log_m(\frac{n}{2l}) \rceil$이며, 최대 노드 수는 $1 + \sum_{i=1}^{h-1}(2 \times m^{i-1}) + \lceil \frac{n}{ml} \rceil$임을 증명하였다.

2.2 알고리즘

CR-트리의 모든 알고리즘에서 순환 도메인에 놓이는 부검색공간은 논리적인 부검색공간으로 변환되어 취급된다. 만약 부검색공간을 구성하는 임의의 폐쇄 경계 구간이 $RI_S > RI_E$인 관계를 갖는다면, 해당 부검색공간은 RI_E에 폐쇄 경계 구간이 속한 도메인의 최대값을 더하여 논리적인 부검색공간으로 변환된다.

CR-트리의 검색 및 삭제 알고리즘은 부검색공간의 취급방법을 제외하고는 R-트리와 동일하기 때문에 본 절에서는 CR-트리의 삽입 알고리즘을 중심으로 기술한다.

객체의 삽입에 따른 전략은 공간국부성 측면에서 매우 중요하다. 객체의 삽입 및 버켓의 분할이 요구될 때, CR-트리의 삽입 알고리즘은 최소의 변형계층분산(modified hierarchical variance) 값을 갖는 색인을 구축함으로써 객체 및 부검색공간의 공간국부성을 최적화한다. 또한 알고리즘에서는 전체 색인의 주기적인 재구축이 요구되지 않는다.

```
Procedure INSERT(O:Object_Type);
{ Input  : An object O
  Output : New CR-Tree index }
VAR
   B       : Bucket_Type;
   BP1,BP2 : Pointer;
BEGIN
   Choose_Bucket(O,BP1);
   Insert_Object(O,B);
   IF B.No_of_Object > MAX_OBJECT THEN (* Bucket overflow *)
      BEGIN
         Assign_New_Bucket(BP2);
         SPLIT_BUCKET(BP1,BP2);
      END;
   Adjust_Tree(B.P_parent);
END;
```

객체의 삽입이 요구되면 CR-트리의 삽입 알고리즘에서는 Choose_Bucket을 사용하여 객체를 포함시켰을 때 전체분산의 증가가 최소가 되는 부검색공간을 경유하여 버켓을 찾고 해당 버켓에 객체를 저장한다.

만약 버켓 및 노드의 범람으로 부검색공간의 분할이 요구된다면, 해당 부검색공간은 SPLIT_BUCKET에 의해 분할 후 생성될 두 개의 부검색공간에 따른 전체분산의 합이 최소가 되도록 분할된다.

버켓의 분할을 위한 알고리즘인 SPLIT_BUCKET은 다음과 같다. 알고리즘에서 Pick_Seeds는 분할 대상 버켓에 존재하는 객체들 중 가장 멀리있는 두 개의 객체를 선택하는 절차이며, Insert_Rest_Objects는 분할 대상 버켓에 남아있는 모든 객체를 새로운 버켓에 저장하는 절차이다. 그리고 Pick_Next는 분할 대상 버켓에 남아있는 임의의 객체를 두개의 새로운 버켓에 각각 포함시켰을 때 버켓에 따른 전체분산 증가분의 차가 가장 큰 객체를 선택하는 절차이다.

```
Procedure SPLIT_BUCKET(BP1,BP2:pointer);
{ Input  : An overflow bucket pointer BP1 and a new bucket
           pointer BP2
  Output : Two buckets after split }
VAR
   B,B1,B2   : Bucket_Type;
   O1,O2     : Object_Type;
BEGIN
   Get_Bucket(BP1,B1); Get_Bucket(BP2,B2);
   B := B1; B1 := null;
   Pick_Seeds(B,O1,O2); (* Pick the first object for each
                           bucket *)
   Insert_Object(O1,B1); Delete_Object(O1,B);
   Insert_Object(O2,B2); Delete_Object(O2,B);
   WHILE B.No_of_Object <> 0 DO
      BEGIN
         IF B1.No_of_Object > (MAX_OBJECT / 2) THEN
            BEGIN
               Insert_Rest_Objects(B.Object,B2); EXIT;
            END;
         IF B2.No_of_Object > (MAX_OBJECT / 2) THEN
            BEGIN
               Insert_Rest_Objects(B.Object,B1); EXIT;
            END;
         Pick_Next(B,B1,B2,O1); (* Select one remaining
                                   object O1 in B *)
         Insert_Object(O1,candidate_bucket(O1,B1,B2));
         Delete_Object(O1,B);
      END;
END;
```

3. 순환 속성을 지원하는 공간색인구조의 성능평가

3.1 실험 환경

실험을 위한 모형은 객체 생성기(object generator), 질의 생성기(query generator), 색인 처리기(index processor), 색인 및 버켓(index and buckets), 해싱표(hashing table), 버퍼(buffer) 등의 구성 요소를 갖는다. 그림 3은 구성 요소들 간의 관계를 보여주고 있다.

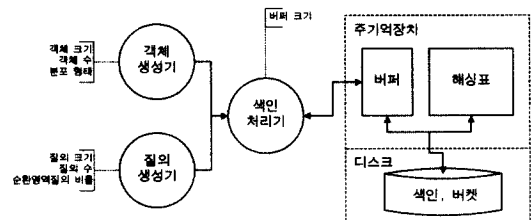



그림 3. 실험 모형

객체 생성기는 객체의 크기, 갯수 그리고 분포 형태를 입력 매개 변수로 하여 객체들을 생성한다. 실험에서는 두 개의 순환 도메인으로 구성된 이차원 검색공간을 가정하였으며, 80바이트의 크기를 갖는 중복되지 않는 20,000개의 객체를 생성하였다. 그림 4는 실험에서 사용된 객체 생성기에 입력되는 객체

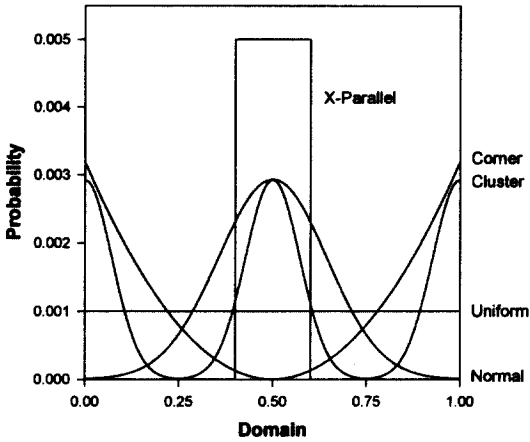



그림 4. 객체의 분포도

의 분포이다. 객체의 분포도는 도메인의 범위 [0,1)내에 발생할 수 있는 1,000개의 객체에 대한 확률을 나타내며, 각 분포에서 확률의 합은 1이다.

균일분포(uniform distribution)는 객체의 분포가 이상적인 상태에서의 색인 처리기의 특성을 규명하기 위한 것이다. 정규분포(normal distribution), 모서리분포(corner distribution), 군집분포(cluster distribution)는 객체들이 검색공간의 일부분에 집중되어 발생되는 분포이다. 정규분포는 객체의 결집시 순환적인 성질이 반영되지 않는 경우, 모서리분포는 순환적 성질이 많이 반영되는 경우 그리고 군집분포는 혼합된 경우에 색인 처리기의 특성을 규명하기 위한 것이다. X-병렬분포(X-parallel distribution)는 검색공간을 구성하는 도메인 중 임의 도메인의 일부분 범위에 객체들이 균일하게 발생되는 경우이다. 이는 임의의 도메인이 제약적인 경우 색인 처리기의 특성을 규명하기 위한 것이다.

질의 생성기는 질의의 크기, 수 그리고 순환 영역질의 비율을 입력 매개 변수로 하여 질의를 생성한다. 실험에서는 다음과 같은 매개 변수 값들을 입력으로 하여 순환 및 선형 영역질의를 균일하게 발생시켰다.

- 영역질의 크기 : 전체 검색공간의 0.01%, 0.1%, 0.5%, 1%
- 영역질의 수 : 1000
- 순환 영역질의 비율 : 0%, 5%, 10%, 30%

순환적인 위치 속성을 갖는 객체의 저장 및 검색에 따른 성능을 평가하기 위해서 색인 처리기는 2장에서 설명된 CR-트리를 사용하였다. 색인 처리기의 입력 매개 변수로 사용한 버퍼의 크기는 1, 2, 4, 8K 바이트이며, 버켓의 이용 정도를 측정하기 위해서 사용한 버퍼의 크기는 1K 바이트부터 31K 바이트까지 이다.

3.2 실험 결과 및 분석

본 절에서는 순환 속성을 지원하는 색인 처리기의 성능을 평가하기 위해 성능 평가 지수로 적중률(hit ratio), 디스크 접근 횟수, 저장공간 사용량, 버켓이용률(bucket utilization) 그리고 색인 구축 및 질의 처리 시간을 고려하였다.

적중률은 주어진 질의에 접근될 것으로 기대되는 객체의 수를 나타내며 다음과 같이 계산된다.

$$\text{적중률} = \frac{\text{발견된 객체의 수}}{\text{버켓용량} \times \text{버켓 접근 횟수}}$$

디스크의 접근 시간은 연산 시간에 비해 매우 크다. 따라서 색인 처리기의 검색 성능을 비교함에 있어 디스크의 접근 횟수는 중요한 성능 평가 지수가 된다.

저장공간 사용량은 색인에 할당된 노드 및 버켓 갯수에 버퍼의 크기를 곱한 것으로 색인 및 모든 버켓을 디스크에 저장하기 위한 바이트 수를 말하며, 버켓이용률은 생성된 전체 버켓에서 실제로 객체가 차지하고 있는 비율을 나타낸다. 버켓이용률은 다음과 같이 계산된다.

$$\text{버켓이용률} = \frac{\text{저장된 객체의 수}}{\text{버켓용량} \times \text{버켓 수}}$$

색인 구축 시간은 색인 처리기가 삽입 연산을 수행하는데 소요되는 모든 처리 시간이다. 즉, 객체를 저장함에 있어 요구되는 노드와 버켓의 접근 및 분할 시간 그리고 색인의 조정 등에 필요한 연산 시간이 고려된다. 질의 처리 시간은 색인 처리기가 질의 처리를 위해 검색 연산을 수행하는데 소요되는 모든 처리 시간이다.

실험결과에 의하면 CR-트리는 균일한 객체의 분포에서보다는 밀집되어 있는 분포에서 상대적으로 더 높은 적중률을 보였다. 그림 5는 버퍼 크기가 1K 바이트이고 영역질의 크기가 0.01%일 때, 균일분포

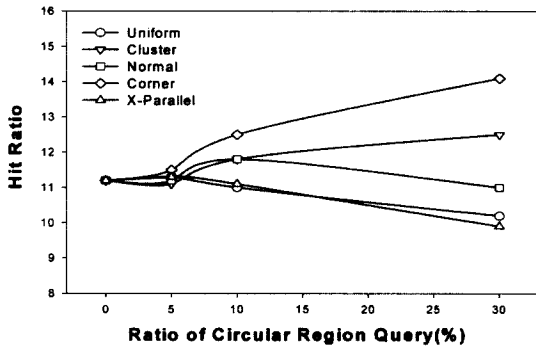



그림 5. 순환 영역질의와 적중률

를 기준으로 하여 순환 영역질의 비율에 따른 각 분포별 적중률 변화를 나타낸 것이다. CR-트리는 순환 영역질의의 비율이 많아질수록 순환적인 성질이 많이 반영된 모서리분포, 군집분포 순으로 적중률이 더욱 높아짐을 보인다. 그러나 순환적인 성질이 반영되어 있지 않은 분포에서는 상대적으로 적중률이 낮아짐을 알 수 있다. 모든 분포에 있어서 CR-트리는 영역질의의 크기가 작고 버퍼의 크기가 클 때 가장 낮은 적중률을 보이며, 영역질의의 크기가 크고 버퍼의 크기가 작을 때 가장 높은 적중률을 보임을 반복된 실험을 통해 알 수 있었다.

영역질의의 크기가 크고 버퍼의 크기가 작을 때 디스크 접근 횟수는 최대가 된다. 본 논문에서는 순환 영역질의 비율과 디스크 접근 횟수의 관계를 최대의 디스크 접근 횟수를 갖는 실험 환경에 맞추어 기술한다.

그림 6은 영역질의의 크기가 1%이고 버퍼의 크기가 1K 바이트일 때, 순환 영역질의의 비율이 높아짐에 따른 디스크 접근 횟수를 각 분포별로 나타낸 것이다. 디스크 접근 횟수는 순환 영역질의 비율이 높아짐에 따라 모서리분포와 군집분포에서는 증가하며, 정규분포와 X-병렬분포에서는 감소한다. 그리고 균일분포에서는 일정함을 보인다. 이러한 이유는 순환적인 특성이 많이 적용된 분포일수록 객체들이 검색공간의 가장자리에 많이 존재함으로써 순환 영역질의에 반응하는 버켓의 수가 많아지기 때문이다.

이상과 같이 CR-트리는 순환적인 위치 속성이 많이 적용된 분포일수록 순환 영역질의의 비율이 높아질 때 디스크 접근 횟수는 증가하지만 적중률은 더욱 높아진다. 이러한 결과는 CR-트리가 색인의 구축시 순환적인 환경과 공간국부성을 고려하여 객체를 결집시키기 때문이다.

그림 7은 버퍼의 크기가 4K 바이트일 때 객체의 삽입에 따른 저장공간 사용량의 증가를 나타낸 것이다. CR-트리의 저장공간 사용량은 모든 분포에서 유사하며, 객체 수가 6,000개일 때 저장공간 사용량의 증가에 변화를 보인다. 이는 객체가 검색공간의 일부분에 집중적으로 삽입됨으로써 해당 검색공간에서 범람이 발생하는 버켓의 수가 많아지기 때문이다.

실험에서는 버켓용량과 버켓이용률 간의 관계를 알아보기 위해 버켓의 크기를 1K 바이트부터 31K 바이트까지 3K 바이트씩 증가시키고, 각 경우에 20,000개의 객체를 삽입한 후 각 분포에 따른 버켓이용률을 산출하였다(그림 8 참조).

실험 결과에 의하면 모든 객체의 분포 형태에서 CR-트리는 버켓용량의 증가에 관계없이 65%에서 70%사이의 변화의 폭이 크지 않는 버켓이용률을 유지하였다. 모서리분포와 군집분포의 경우에서도 이러한 버켓이용율이 유지되는 이유는, 순환 도메인이 포함된 검색공간에서 CR-트리는 순환적인 환경을

그림 6. 순환 영역질의 비율에 따른 디스크 접근 횟수

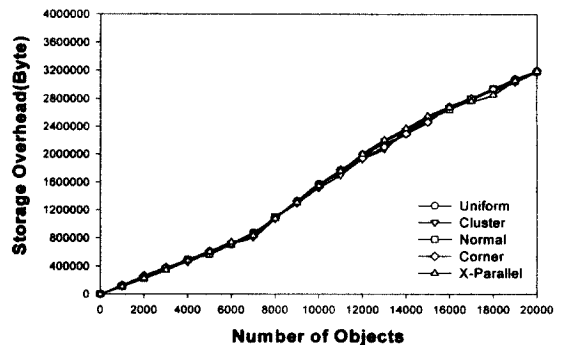



그림 7. 객체 삽입에 따른 저장공간 사용량

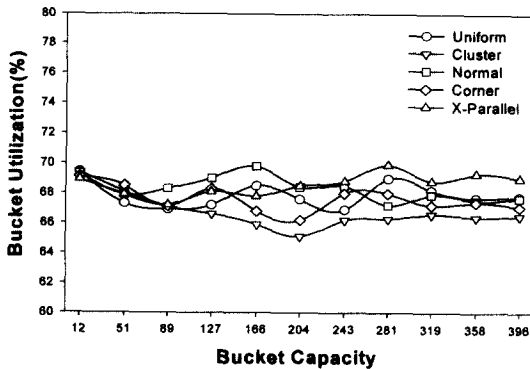



그림 8. 버켓용량에 따른 버켓 이용률

전제로 각 버켓에 저장되는 객체의 수를 고려하여 객체를 결집시킴으로써 버켓의 수를 줄였기 때문이다.

그림 9는 색인 구축 및 질의 처리를 위한 CR-트리의 연산 시간을 나타낸 것이다. 이를 위해 색인 처리기의 연산 횟수가 가장 많이 요구되는 1K 바이트 크기의 버퍼를 입력 매개 변수로 선택하였다. 그리고 색인 처리기의 질의 처리 시간을 분석하기 위해, 실험에서는 영역질의 크기(0.01%, 0.1%, 0.5%, 1%)와 순환 영역질의 비율(0%, 5%, 10%, 30%)에 따른 16가지의 경우를 고려하여 20,000개의 영역질의를 균일하게 생성한 후 처리 시간을 측정하였다.

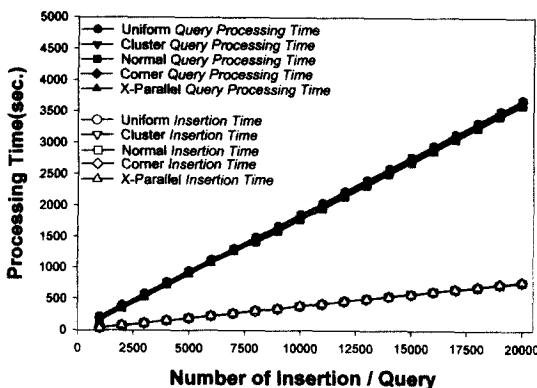



그림 9. 색인 구축 시간과 질의 처리 시간

그림에서 하단의 그래프들은 삽입되는 객체의 갯수에 대한 누적된 색인 구축 시간의 변화를, 상단의 그래프들은 영역질의에 대한 누적된 질의 처리 시간의 변화를 각 분포별로 보여준다. 색인 구축 시간과 질의 처리 시간은 삽입되는 객체의 갯수가 많아질수록 그리고 질의의 갯수가 증가할수록 선형적으로 커지며 객체의 분포에 관계없이 동일함을 보였다. 이러한 결과는 CR-트리의 삽입 알고리즘에 순환적인 객체의 분포 환경에서도 공간국부성을 최적화하기 위한 전략이 포함되어있기 때문이다.

4. 결 론

공간 데이타베이스 시스템의 성능을 향상시키기 위해서는 동적 및 정적 환경에서 발생하는 공간객체를 효율적으로 관리할 수 있는 공간색인구조가 필요하다. 그리고 검색의 성능을 높이기 위해서는 객체들의 위치 속성을 고려한 공간색인구조가 요구되며, 특히 객체가 가질 수 있는 순환적인 위치 속성을 고려함으로써 검색의 성능을 한 단계 더 향상시킬 수 있다.

본 논문에서는 객체의 순환적인 위치 속성을 지원하는 공간색인구조의 성능을 실험을 통해 평가하였다. 이를 위해 색인 처리기로는 CR-트리를 이용하였다. CR-트리는 순환적인 성질을 갖는 도메인이 포함된 검색공간에서 공간국부성을 고려하여 객체를 결집시키는 공간색인구조이다.

실험에서는 CR-트리의 성능을 적중률, 버켓이용률 그리고 색인 구축 및 질의 처리 시간에 따라 평가하였다. 실험 데이타로는 균일분포, 정규분포, 모서리분포, 군집분포, X-병렬분포를 이루는 객체들을 이용하였다.

반복된 실험결과에 의하면, CR-트리는 순환적으로 객체의 결집이 가능한 분포에서 순환 영역질의의 비율이 커질수록 적중률이 높아졌다. 또한 모든 객체의 분포 형태에서 버켓용량의 증가에 관계없이 65%에서 70%사이의 변화의 폭이 크지 않는 버켓이용률을 유지하였다. 그리고 CR-트리의 저장공간 사용량과 색인 구축 및 질의 처리 시간은 객체의 분포에 관계없이 동일함을 보였다. 이러한 결과는 CR-트리의 알고리즘에 순환적인 객체의 분포 환경에서도 공간국부성을 최적화하기 위한 전략이 포함되어있기 때문이다.

따라서 순환도메인이 포함된 다차원 검색공간이 존재하고 객체의 분포가 순환적 위치 속성이 많이 반영되어 있을수록, 순환적인 위치 속성을 고려하여 객체를 결집시킴으로써 적중률 및 버켓이용률을 향상시킬 수 있다.

참 고 문 헌

[1] W.G.Aref and H.Samet, "Optimization Strategies for Spatial Query Processing," Proc. of the 7th Int. Conf. on VLDB, pp.81-90, 1991.

[2] A.Biliris, "An Efficient Database Storage Structure for Large Dynamic Objects," Proc. of the 8th Int. Conf. on Data Engineering, pp.301-308, 1992.

[3] T.Brinkhoff and H.P.Kriegel, "The Impact of Global Clustering on Spatial Database Systems," Proc. of the 20th VLDB Conf. pp.168-179, 1994.

[4] O.Guenther and A.Buchmann, "Research Issues in Spatial Databases," ACM SIGMOD Record, Vol.19, No.4, pp.61-68, 1990.

[5] O.Günther, "Efficient Computation of Spatial Joins," Proc. of the 9th Int. Conf. on Data Engineering, pp.50-59, 1993.

[6] L.K.Joune and L.Robert, "The Spatial Locality and a Spatial Indexing Method by Dynamic Clustering in Hypermap System," Proc. of the 2nd Sym. on Large Spatial Databases, pp.207-223, 1991.

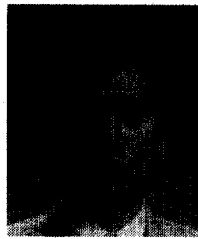
[7] H.Lu, B.C.Ooi, A.Souza and C.C.Low, "Storage Management in Geographic Information Systems," Proc. of 2nd Sym. on Large Spatial Databases, pp.451-470, 1991.

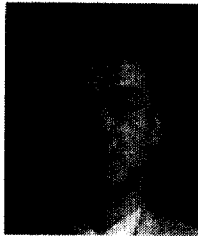
[8] C.B.Mederios and F.Pires, "Databases for GIS," Proc. ACM SIGMOD Int. Conf. on Management of Data, Vol.23, No.1, pp.107-115, 1994.

[9] N.Roussopoulos, S.Kelley and F.Vincent, "Nearest Neighbor Queries," Proc. ACM SIGMOD Int. Conf. on Management of Data, pp.71-79, 1995.

[10] 김홍기, "공간 데이타베이스 시스템에서 객체의 동적 결집을 위한 공간색인구조," 전남대학교 박사학위 논문, 1996.

[11] 김홍기, 황부현, "순환도메인을 기반으로 하는 PR-화일의 구현 및 성능 평가," 한국정보처리학회 논문지, 3권 1호, pp.63-76, 1996. 1.

김 홍 기

1984년 전남대학교 계산통계학과(이학사)
1986년 전남대학교 대학원 계산통계학과(이학석사)
1996년 전남대학교 대학원 전산통계학과(이학박사)
1991년~현재 동신대학교 컴퓨터학과 부교수
관심분야 : 공간데이타베이스, 컴퓨터그래픽스, 멀티미디어시스템

선 희 준

1988년 목포대학교 전산통계학과(이학사)
1990년 전남대학교 대학원 전산통계학과(이학석사)
1998년 전남대학교 대학원 전산통계학과(이학박사)
1997년~현재 서남대학교 컴퓨터영상·정보통신학과 조교수
관심분야 : 공간데이타베이스, 멀티미디어시스템, 지리정보시스템